



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Transmit Patent Application of:)
ANSORGE ET AL.)
Serial No. 10/622,021)
Confirmation No: 1685)
Filing Date: JULY 17, 2003)
For: METHOD AND DEVICE FOR ENCODING)
WIDEBAND SPEECH)

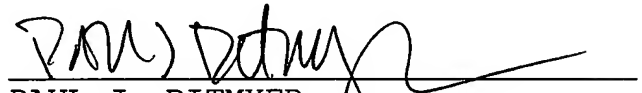
TRANSMITTAL OF CERTIFIED PRIORITY DOCUMENT

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

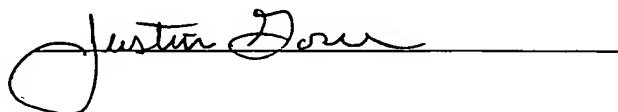
Transmitted herewith is a certified copy of the
priority European Application No. 02015918.2.

Respectfully submitted,


PAUL J. DITMYER
Reg. No. 40,455
Allen, Dyer, Doppelt, Milbrath
& Gilchrist, P.A.
255 S. Orange Avenue, Suite 1401
Post Office Box 3791
Orlando, Florida 32802
Telephone: 407/841-2330
Fax: 407/841-2343
Attorney for Applicant

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being
deposited with the United States Postal Service as first class
mail in an envelope addressed to: COMMISSIONER FOR PATENTS,
P.O. BOX 1450, ALEXANDRIA, VA 22313-1450, on this 16th day of
November, 2004.





**Europäisches
Patentamt**

**European
Patent Office**

**Office européen
des brevets**

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02015918.2

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk



Anmeldung Nr:
Application no.: 02015918.2
Demande no:

Anmeldetag:
Date of filing: 17.07.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

STMicroelectronics N.V.
Tower B, 17th Floor,
World Trade Center,
Strawinskylaan 1725
1077 XX Amsterdam
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se référer à la description.)

Procédé et dispositif d'encodage de la parole à bande élargie

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

G10L/

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR

B 02/1704 EP

Société dite : STMicroelectronics N.V.

Procédé et dispositif d'encodage de la parole à bande élargie.

Procédé et dispositif d'encodage de la parole à bande élargie.

5 L'invention concerne l'encodage/décodage de la parole à bande élargie, notamment mais non limitativement pour la téléphonie mobile.

En bande élargie, la bande passante du signal de parole est comprise entre 50 et 7000 Hz.

10 Des séquences de parole successives échantillonnées à une fréquence d'échantillonnage prédéterminée, par exemple 16 kHz, sont traitées dans un dispositif de codage utilisant une prédiction linéaire à excitation par séquences codées (ACELP : « algebraic-code-excited linear-prediction »), bien connu de l'homme du métier, et décrit notamment dans la recommandation ITU-TG 729, version 3/96, intitulée « codage de la parole à 8 kbits/s par prédiction
15 linéaire avec excitation par séquences codées à structure algébrique conjuguée ».

On va maintenant rappeler brièvement, en se référant à la figure 1, les principales caractéristiques et fonctionnalités d'un tel codeur, l'homme du métier pouvant se référer à toutes fins utiles,
20 pour plus de détails, à la recommandation G 729 précitée.

Le codeur de prédiction CD, du type ACELP, est fondé sur le modèle de codage prédictif linéaire à excitation par code. Le codeur opère sur des super-trames vocales équivalentes par exemple à 20 ms de signal et comportant chacune 320 échantillons. L'extraction des
25 paramètres de prédiction linéaire, c'est-à-dire les coefficients du filtre de prédiction linéaire également appelé filtre de synthèse à court-terme $1/A(z)$, est effectuée pour chaque super-trame de parole. Par contre, chaque super-trame est subdivisée en trames de 5 ms comprenant 80 échantillons. Toutes les trames, le signal vocal est
30 analysé pour en extraire les paramètres du modèle de prédiction CELP (c'est-à-dire, notamment, un mot numérique d'excitation à long terme v_i extrait d'un répertoire codé adaptatif DLT, également dénommé « dictionnaire à long terme adaptatif », un gain à long terme associé G_a , un mot d'excitation à court terme c_j , extrait d'un

répertoire codé algébrique DCT, également dénommé « répertoire codé fixe » ou « dictionnaire à court terme algébrique », et un gain à court terme associé G_c).

Ces paramètres sont ensuite codés et transmis.

5 A la réception, ces paramètres servent, dans un décodeur, à récupérer les paramètres d'excitation et du filtre prédictif. On reconstitue alors la parole en filtrant ce flux d'excitation dans un filtre de synthèse à court terme.

10 Alors que le dictionnaire adaptatif DLT contient des mots numériques représentatifs de délais tonaux représentatifs d'excitations passées, le dictionnaire à court terme DCT est fondé sur une structure algébrique utilisant un modèle de permutation entrelacée d'impulsions de Dirac. Dans ce répertoire codé, qui
15 contient des excitations innovatrices également appelées excitations algébriques ou à court-terme, chaque vecteur contient un certain nombre d'impulsions non nulles, par exemple quatre, dont chacune peut avoir l'amplitude +1 ou -1 avec des positions prédéterminées.

20 Les moyens de traitement du codeur CD comportent fonctionnellement des premiers moyens d'extraction MEXT1 destinés à extraire le mot d'excitation à long terme, et des deuxièmes moyens d'extraction MEXT2 destinés à extraire le mot d'excitation à court terme. Fonctionnellement, ces moyens sont réalisés par exemple de façon logicielle au sein d'un processeur.

25 Ces moyens d'extraction comportent un filtre prédictif FP ayant une fonction de transfert égale à $1/A(z)$, ainsi qu'un filtre de pondération perceptuel FPP ayant une fonction de transfert $W(z)$. Le filtre de pondération perceptuelle est appliqué au signal pour modéliser la perception de l'oreille.

30 Par ailleurs, les moyens d'extraction comportent des moyens MECM destinés à effectuer une minimisation d'une erreur carrée moyenne.

Le filtre de synthèse FP de la prédiction linéaire modélise l'enveloppe spectrale du signal. L'analyse prédictive linéaire est effectuée toutes les super-frames, de façon à déterminer les

coefficients de filtrage prédictif linéaire. Ceux-ci sont convertis en paires de lignes spectrales (LSP : « Line Spectrum Pairs ») et numérisés par quantification vectorielle prédictive en deux étapes.

5 Chaque super-trame de parole de 20 ms est divisée en quatre trames de 5 ms chacune contenant 80 échantillons. Les paramètres LSP quantifiés sont transmis au décodeur une fois par super-trame alors que les paramètres à long terme et à court terme sont transmis à chaque trame.

10 Les coefficients du filtre de prédiction linéaire, quantifiés et non quantifiés, sont utilisés pour la trame la plus récente d'une super-trame, tandis que les trois autres trames de la même super-trame utilisent une interpolation de ces coefficients. Le délai tonal en boucle ouverte est estimé toutes les deux trames sur la base du signal vocal pondéré perceptuellement. Puis, les opérations suivantes
15 sont répétées à chaque trame :

Le signal cible à long terme X_{LT} est calculé en filtrant le signal de parole échantillonné $s(n)$ par le filtre de pondération perceptuelle FPP.

20 On soustrait ensuite du signal vocal pondéré la réponse à entrée nulle du filtre de synthèse pondéré FP, FPP, de façon à obtenir un nouveau signal cible long terme.

La réponse impulsionnelle du filtre de synthèse pondéré est calculé.

25 Une analyse tonale en boucle fermée utilisant une minimisation de l'erreur carrée moyenne, est ensuite effectuée afin de déterminer le mot d'excitation à long terme v_i et le gain associé G_a , au moyen du signal cible et de la réponse impulsionnelle, par recherche autour de la valeur du délai tonal en boucle ouverte.

30 Le signal cible long terme est ensuite mis à jour par soustraction de la contribution filtrée y du répertoire codé adaptatif DLT et ce nouveau signal cible court terme X_{ST} est utilisé lors de l'exploration du répertoire codé fixe DCT afin de déterminer le mot d'excitation court terme c_j et le gain G_c associé. Là encore, cette

recherche en boucle fermée s'effectue par une minimisation de l'erreur carrée moyenne.

5 Finalement, le dictionnaire à long terme adaptatif DLT ainsi que les mémoires des filtres FP et FPP, sont mis à jour au moyen des mots d'excitation long terme et court terme ainsi déterminés.

10 La qualité d'un algorithme CELP dépend fortement de la richesse du dictionnaire d'excitation algébrique DCT. Si l'efficacité d'un tel algorithme est incontestable pour les signaux à bande passante étroite (300-3400 Hz), des problèmes surviennent pour des signaux à bande élargie.

 Les inventeurs ont en effet observé que même avec un dictionnaire algébrique très riche, l'algorithme d'encodage de la parole produit deux types de problèmes :

- 15 1) une qualité globale de parole reconstruite totalement insuffisante (la parole reconstruite manque de présence, le niveau d'énergie est très variable, le timbre de la voix est méconnaissable, ...),
- 20 2) un signal reconstruit corrompu par trois sortes de bruits :
 - un bruit harmonique en haute fréquence (comb-like noise),
 - un fort bruit à haute fréquence, tel un bruit de quantification,
 - un bruit en basse fréquence (rumbling noise), tel qu'un balai de paille frappé à intervalles réguliers sur le sol.

25 Il a alors été observé qu'une amélioration de la qualité globale de la parole pouvait être obtenue par une élimination partielle ou totale de ces bruits.

 Un but de l'invention est de réduire le bruit harmonique et le bruit haute fréquence.

30 L'invention a également pour but de supprimer le bruit de type « sifflement » entachant les trames de parole voisée.

 L'invention a encore pour but de contrôler indépendamment les distorsions à court terme et à long terme.

L'invention propose donc un procédé d'encodage de la parole à bande élargie, dans lequel on échantillonne la parole de façon à obtenir des trames vocales successives comportant chacune un nombre prédéterminé d'échantillons, et à chaque trame vocale, on détermine des paramètres d'un modèle de prédiction linéaire à excitation par code, ces paramètres comportant un mot numérique d'excitation à long terme extrait d'un répertoire codé adaptatif et un gain à long terme associé, ainsi qu'un mot d'excitation à court terme extrait d'un répertoire codé algébrique et un gain à court terme associé, et on met à jour le répertoire codé adaptatif à partir du mot d'excitation à long terme extrait et du mot d'excitation à court terme extrait.

Selon une caractéristique générale de l'invention, on somme le produit du mot extrait d'excitation à long terme par le gain à long terme associé, avec le produit du mot extrait d'excitation à court terme par le gain à court terme associé, on filtre le mot numérique sommé dans un filtre passe-bas ayant une fréquence de coupure supérieure au quart de la fréquence d'échantillonnage et inférieure à la moitié de celle-ci, et on met à jour le répertoire codé adaptatif avec le mot filtré.

L'invention utilise ici un filtre « de correction totale » qui combine un filtre de correction du bruit harmonique et un filtre de correction haute fréquence.

L'invention permet ainsi une amélioration de la qualité durant les trames de parole voisées. Par ailleurs, la complexité de l'encodeur est réduite en fusionnant en un seul filtre le filtre de correction harmonique et le filtre de correction haute fréquence.

L'invention se distingue notamment d'une solution décrite dans un article de Kroon et Atal, intitulé « Strategies for Improving the Performance of CELP Coders at Low Bit Rates », Proc., IEEE, Int. Conf. Acoustics, Speech, and Signal Processing, ICASSP'88, New York, USA, 1988, Pages 151-154, qui propose un filtrage du dictionnaire adaptatif effectué à la sortie de ce dictionnaire et non pas à l'entrée conformément à l'invention.

Ainsi, le préfiltrage du dictionnaire adaptatif selon l'invention présente par rapport au post-filtrage de l'article de Kroon et Atal, l'avantage que le filtrage est pris en compte lors de la minimisation de l'erreur effectuée pour choisir l'excitation adaptative à la trame suivante. Ce n'est pas le cas pour la solution de Kroon et Atal, puisque le filtrage proposé a lieu sur l'excitation choisie suivante. Aussi, pour tenir compte du filtrage dans la minimisation de l'erreur, il faudrait alors augmenter considérablement la complexité et filtrer toute excitation à tester.

Selon un mode de mise en œuvre préférentielle, on filtre le mot sommé avec un filtre numérique à réponse impulsionnelle finie à phase linéaire ayant un ordre au moins égal à 10. Par exemple, lorsque la fréquence d'échantillonnage est de 16 kHz, on peut choisir un filtre d'ordre 20 ayant une fréquence de coupure de l'ordre de 6 kHz.

Bien que la qualité de la parole soit ainsi améliorée, les trames de parole voisées semblent encore corrompues par un bruit de type « sifflement ». Ce bruit de nature haute fréquence provient de l'excitation à court terme qui introduit des artéfacts indésirables. Deux types de solutions pour résoudre ce problème ont déjà été proposés dans la littérature. Un premier type de solution, décrit par exemple

dans l'article de Gerson et Jasiuk, intitulé « Techniques for Improving the Performance of CELP-Type Speech Coders », IEEE, Journal on Selected Areas In Communications, Vol. 10, N°5, Juin 1992, pages 858-865, ou bien

dans l'article de Miki et autres, intitulé « A Pitch Synchronous Innovation CELP (PSI-CELP) Coder for 2-4 kbit/s », Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, ICASSP'94, Adelaïde, South Australia, 1994, Vol. II, pages 113-116,

propose de rendre la contribution à court terme périodique.

Un autre type de solution, décrit par exemple

dans l'article de Taniguchi Johnson et Ohta, intitulé « Pitch Sharpening for Perceptually Improved CELP, and the Sparse-Delta Codebook for Reduced Computation », Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, ICASSP'91, Toronto, Canada, 1991, pages 241-244, ou

dans l'article de Shoham, intitulé « Constrained-Stochastic Excitation Coding of Speech At 4,8 kb/s », Advances in Speech Coding, B.S. Atal, V. Cuperman, and A. Gersho, Eds., Dordrecht, The Netherlands, Kluwer, 1991, pages 339-348,

propose un contrôle de façon adaptative du gain à court terme.

L'invention propose également une solution du type contrôle de gain, mais totalement différente de celle décrite notamment dans les articles de Taniguchi et autres et de Shoham.

Plus précisément, selon un mode de mise en œuvre de l'invention, l'extraction du mot d'excitation à court terme comporte un filtrage numérique de prédiction linéaire et le procédé comporte une mise à jour de l'état du filtre de prédiction linéaire avec le mot d'excitation à court terme filtré par un filtre dont le ou les coefficients dépendent de la valeur du gain à long terme, de façon à affaiblir la contribution de l'excitation à court terme lorsque le gain de l'excitation à long terme est supérieur à un seuil prédéterminé, par exemple égal à 1.

En d'autres termes, la solution selon l'invention consiste ici à affaiblir la contribution de l'excitation à court terme si le gain de l'excitation à long terme est important. Cependant, c'est la contribution de l'excitation à court terme non affaiblie qui est stockée dans le dictionnaire adaptatif pour sa mise à jour. Ainsi, la réduction intervient seulement sur la sortie. Le fait de préserver l'amplitude de la contribution à court terme à stocker est important, puisque la richesse du dictionnaire adaptatif est ainsi conservée pour les plus basses fréquences.

Bien entendu, la correction du gain doit également être appliquée lors de la reconstruction du signal au niveau du décodeur.

Ce filtre peut être d'ordre 0 ou bien d'ordre supérieur ou égal à 1. Dans ce dernier cas, le filtre d'ordre supérieur ou égal à 1 peut être à réponse impulsionnelle finie.

5 Selon un mode de mise en œuvre de l'invention, dans lequel filtre est d'ordre 1 et sa fonction de transfert égale à $B_0 + B_1 z^{-1}$, le premier coefficient B_0 du filtre est égal à $1/(1 + \beta \cdot \min(Ga, 1))$, et le deuxième coefficient B_1 du filtre est égal à $\beta \cdot \min(Ga, 1)/(1 + \beta \cdot \min(Ga, 1))$, où β est un nombre réel de valeur absolue inférieure à 1, Ga est le gain à long terme et $\min(Ga, 1)$ désigne la valeur minimale entre Ga et 1.

10 Selon une autre variante de l'invention, qui peut être prise en combinaison ou bien indépendamment de la variante précédente, on effectue l'extraction du mot d'excitation à long terme en utilisant un premier filtre de pondération perceptuelle comportant un premier
15 filtre de pondération formantique, et on effectue l'extraction du mot d'excitation à court terme en utilisant le premier filtre de pondération perceptuelle cascadié à un deuxième filtre de pondération perceptuelle comportant un deuxième filtre de pondération formantique. Le dénominateur de la fonction de transfert
20 du premier filtre de pondération formantique est égal au numérateur du deuxième filtre de pondération formantique.

Ainsi, selon cette variante, l'utilisation de deux filtres de pondération formantique différents permet de contrôler indépendamment les distorsions à court terme et à long terme. Le
25 filtre de pondération à court terme est cascadié au filtre de pondération à long terme. En outre, le fait de lier le dénominateur du filtre de pondération à long terme au numérateur du filtre de pondération à court terme permet de contrôler séparément ces deux filtres et permet en outre une nette simplification lorsque ces deux
30 filtres sont cascadiés.

Bien entendu, lorsque cette variante est utilisée en combinaison avec la variante de contrôle de gain, il est prévu une mise à jour de l'état des deux filtres de pondération perceptuelle

avec le mot d'excitation à court terme filtré par le filtre d'ordre supérieur ou égal à 1.

L'invention a également pour objet un dispositif d'encodage de la parole à bande élargie, comportant

5 -des moyens d'échantillonnage aptes à échantillonner la parole de façon à obtenir des trames vocales successives comportant chacune un nombre prédéterminé d'échantillons,

 -des moyens de traitement aptes à chaque trame vocale, à déterminer des paramètres d'un modèle de prédiction linéaire à excitation par code, ces moyens de traitement comportant des premiers moyens d'extraction aptes à extraire un mot numérique d'excitation à long terme d'un répertoire codé adaptatif et à calculer un gain à long terme associé, et des deuxièmes moyens d'extraction aptes à extraire un mot d'excitation à court terme d'un répertoire codé algébrique et à calculer un gain à court terme associé, et

10 -des premiers moyens de mise à jour aptes à mettre à jour le répertoire codé adaptatif à partir du mot d'excitation à long terme extrait et du mot d'excitation à court terme extrait. Selon une caractéristique générale de l'invention, les premiers moyens de mise

15 à jour comportent

 -des premiers moyens de calcul aptes à sommer le produit du mot extrait d'excitation à long terme par le gain à long terme associé, avec le produit du mot extrait d'excitation à court terme par le gain à court terme associé, de façon à délivrer un mot numérique sommé, et

20 -un filtre passe-bas ayant une fréquence de coupure supérieure au quart de la fréquence d'échantillonnage et inférieure à la moitié de celle-ci, et connecté entre la sortie des premiers moyens de calcul et le répertoire codé adaptatif de façon à mettre à jour ce

25 répertoire adaptatif avec le mot filtré.

30 Selon un mode de réalisation de l'invention, les premiers moyens d'extraction comportent un filtre numérique de prédiction linéaire, et par le fait que le dispositif comporte des deuxièmes moyens de mise à jour aptes à effectuer une mise à jour de l'état du

5 filtre de prédiction linéaire avec le mot d'excitation à court terme filtré par un filtre dont le ou les coefficients dépendent de la valeur du gain à long terme, de façon à affaiblir la contribution de l'excitation à court terme lorsque le gain de l'excitation à long terme est supérieur à un seuil prédéterminé.

10 Selon un autre mode de réalisation de l'invention, les premiers moyens d'extraction comprennent un premier filtre de pondération perceptuelle comportant un premier filtre de pondération formantique, par le fait que les deuxièmes moyens d'extraction comprennent le premier filtre de pondération perceptuelle cascadié à un deuxième filtre de pondération perceptuelle comportant un deuxième filtre de pondération formantique, et par le fait que le dénominateur de la fonction de transfert du premier filtre de pondération formantique est égal au numérateur du deuxième filtre de pondération formantique.

15 L'invention a également pour objet un terminal d'un système de communication sans fil, par exemple un téléphone mobile cellulaire, incorporant un dispositif tel que défini ci-avant.

20 D'autres avantages et caractéristiques de l'invention apparaîtront à l'examen de la description détaillée de modes de réalisation et de mise en œuvre, nullement limitatifs, et des dessins annexés, sur lesquels :

-la figure 1, déjà décrite, illustre schématiquement un dispositif d'encodage de la parole, selon l'art antérieur ;

25 -la figure 2 illustre schématiquement un premier mode de réalisation d'un dispositif d'encodage, selon l'invention ;

-la figure 3 illustre schématiquement un deuxième mode de réalisation d'un dispositif d'encodage, selon l'invention, et la figure 3a illustre schématiquement un mode de réalisation d'un décodeur correspondant ;

30 -la figure 4 illustre schématiquement un troisième mode de réalisation d'un dispositif d'encodage, selon l'invention ;

-la figure 5 illustre schématiquement un quatrième mode de réalisation d'un dispositif d'encodage, selon l'invention ; et

-la figure 6 illustre schématiquement l'architecture interne d'un téléphone mobile cellulaire incorporant un dispositif de codage, selon l'invention.

5 Le dispositif d'encodage, ou codeur, CD, selon l'invention, tel qu'illustré sur la figure 2, se distingue de celui de l'art antérieur comme illustré sur la figure 1 par le fait que les moyens MAJ de mise à jour du dictionnaire à long terme adaptatifs DLT comportent un filtre de correction totale FLCT connecté entre la sortie d'un sommateur SM et l'entrée du dictionnaire DLT.

10 Les deux entrées du sommateur SM reçoivent respectivement le produit du mot extrait d'excitation à long terme v_i par le gain à long terme associé G_a , et le produit du mot extrait d'excitation à court terme c_j par le gain associé G_c .

15 Ce filtre de correction totale FLCT est un filtre passe-bas ayant d'une façon générale une fréquence de coupure supérieure au quart de la fréquence d'échantillonnage et inférieure à la moitié de celle-ci.

20 Ce filtre est dans l'exemple décrit un filtre numérique à réponse impulsionnelle finie à phase linéaire ayant un ordre au moins égal à 10.

25 Plus précisément, lorsque la fréquence d'échantillonnage est de 16 kHz, on utilisera de préférence une fréquence de coupure de l'ordre de 6 kHz et un filtre d'ordre 20, ce qui réalise un bon compromis entre la complexité de la mémoire et la qualité du signal vocal reconstitué.

30 Le bruit harmonique est introduit par la contribution de l'excitation à long terme et par la répétition d'échantillons pour des valeurs de la période fondamentale (pitch) inférieures à la longueur d'une trame de parole, ici de 5 ms. Ce bruit est également présent pour des valeurs de la période fondamentale supérieures à la taille d'une trame. Il est en outre lié au gain adaptatif, extrait une seule fois par trame de parole.

L'utilisation d'un filtrage passe-bas de la contribution à long terme est une solution pour réduire le bruit harmonique.

Par ailleurs, le bruit à haute fréquence est introduit par des anciennes contributions à haute fréquence du dictionnaire à court terme, présentes dans le dictionnaire adaptatif.

5 Pour éliminer ce bruit haute fréquence, on peut éliminer les composantes résiduelles à haute fréquence du dictionnaire adaptatif, en utilisant un filtre de correction, et ceci avant de remettre à jour le dictionnaire.

10 Le filtre de correction totale selon l'invention réalise donc la double fonction de correction harmonique et de correction haute fréquence. Ceci permet une amélioration de la qualité durant les trames de parole voisées.

15 En outre, l'emplacement de ce filtre, c'est-à-dire en entrée du dictionnaire adaptatif, permet de prendre en compte le filtrage lors de la minimisation de l'erreur effectuée pour choisir l'excitation adaptative de la trame de parole suivante.

20 Dans le mode de réalisation illustré sur la figure 3, le codeur CD comporte en outre des deuxièmes moyens de mise à jour MAJ2 aptes à effectuer une mise à jour de l'état du filtre de prédiction linéaire FP et de l'état du filtre de pondération perceptuelle FPP avec le mot d'excitation à court terme c_j filtré par un filtre que l'on a représenté ici schématiquement par un gain Gc' . Ce filtre peut être d'ordre 0 et son gain Gc' est inférieur au gain Gc . En variante, ce filtre peut être à réponse impulsionnelle finie et d'ordre supérieur ou égal à 1, avec pour cas particulier un filtre à réponse impulsionnelle finie d'ordre 1.

25 Les coefficients de ce filtre d'ordre 1 dépendent de la valeur du gain à long terme G_a , de façon à affaiblir la contribution de l'excitation à court terme lorsque le gain de l'excitation à long terme G_a est supérieur à un seuil prédéterminé, par exemple égal à 1.

30 La fonction de transfert de ce filtre est égale à $B_0 + B_1 z^{-1}$. A titre d'exemple, le premier coefficient du filtre B_0 peut être déterminé par la formule (I) ci-dessous.

$$1/(1 + 0,98 \min (G_a, 1)) \quad (I)$$

tandis que le deuxième coefficient du filtre B1 peut être déterminé par la formule (II) ci-dessous.

$$5 \quad 0,98 \min (Ga, 1) / (1 + 0,98 \min (Ga, 1)) \quad (II)$$

Par contre c'est bien la contribution à court terme non affaiblie (gain Gc) qui est stockée dans le dictionnaire adaptatif DLT pour sa mise à jour.

10 Ainsi, l'affaiblissement intervient uniquement sur le signal de sortie et le fait de préserver l'amplitude de la contribution à court terme à stocker permet de conserver la richesse du dictionnaire adaptatif pour les plus basses fréquences.

15 Naturellement, la correction du gain Gc doit être également appliquée pour la mise à jour de l'état des mémoires des filtres dans le décodeur DCD, comme illustré schématiquement sur la figure 3a.

La variante de réalisation illustrée sur la figure 3 permet, outre les avantages apportés par le filtre de correction totale, d'éliminer le bruit de type sifflement sur les trames de parole voisées.

20 Le filtre de pondération perceptuelle FPP exploite les propriétés de masquage de l'oreille humaine par rapport à l'enveloppe spectrale du signal de parole, dont la forme est fonction des résonances du conduit vocal. Ce filtre permet d'attribuer plus d'importance à l'erreur apparaissant dans les vallées spectrales par rapport aux pics formantiques.

25 Dans les variantes illustrées sur les figures 2 et 3, le même filtre de pondération perceptuelle FPP est utilisé pour la recherche à court terme et pour celle à long terme. La fonction de transfert W(z) de ce filtre FPP est donnée par la formule (III) ci-dessous.

$$30 \quad W(z) = \frac{A(z/\gamma_1)}{A(z/\gamma_2)} \quad (III)$$

dans laquelle $1/A(z)$ est la fonction de transfert du filtre prédictif FP et γ_1 et γ_2 sont les coefficients de pondération perceptuelle, les deux coefficients étant positifs ou nuls et inférieurs ou égaux à 1 avec le coefficient γ_2 inférieur ou égal au coefficient γ_1 .

5 D'une façon générale, le filtre de pondération perceptuelle est constitué d'un filtre de pondération formantique et d'un filtre de pondération de la pente de l'enveloppe spectrale du signal (tilt).

Dans le cas présent, on supposera que le filtre de pondération perceptuelle est uniquement formé du filtre de pondération formantique dont la fonction de transfert est donnée par la formule (III) ci-dessus.

Or, la nature spectrale de la contribution à long terme est différente de celle de la contribution à court terme. Par conséquent, il est avantageux d'utiliser deux filtres de pondération formantique différents, permettant de contrôler indépendamment les distorsions à court terme et à long terme.

Un tel mode de réalisation est illustré sur la figure 4, dans laquelle, par rapport à la figure 3, le filtre unique FPP a été remplacé par un premier filtre de pondération formantique FPP1 pour la recherche à long terme, cascadié avec un deuxième filtre de pondération formantique FPP2 pour la recherche à court terme.

Puisque le filtre de pondération à court terme FPP2 est cascadié au filtre de pondération à long terme, les filtres apparaissant dans la boucle de recherche long terme doivent aussi apparaître dans la boucle de recherche court terme.

La fonction de transfert $W_1(z)$ du filtre de pondération formantique FPP1 est donnée par la formule (IV) ci-dessous.

$$W_1(z) = \frac{A(z/\gamma_{11})}{A(z/\gamma_{12})} \quad (IV)$$

30

tandis que la fonction de transfert $W_2(z)$ du filtre de pondération formantique FPP2 est donnée par la formule (V) ci-dessous.

$$W_2(z) = \frac{A(z/\gamma_{21})}{A(z/\gamma_{22})} \quad (V)$$

Par ailleurs, le coefficient γ_{12} est égal au coefficient γ_{21} . Ceci permet une nette simplification lorsqu'on cascade ces deux filtres. Ainsi, le filtre équivalent à la cascade de ces deux filtres a une fonction de transfert donnée par la formule (VI) ci-dessous.

$$\frac{A(z/\gamma_{11})}{A(z/\gamma_{22})} \quad (VI)$$

Par ailleurs, si l'on utilise la valeur 1 pour le coefficient γ_{11} , alors le filtre de synthèse FP (ayant la fonction de transfert $1/A(z)$) suivi du filtre de pondération à long terme FPP1 et du filtre de pondération FPP2 équivaut alors au filtre dont la fonction de transfert est donnée par la formule (VII) ci-dessous.

$$\frac{1}{A(z/\gamma_{22})} \quad (VII)$$

Ce qui réduit encore considérablement la complexité de l'algorithme d'extraction des excitations.

A titre indicatif, on peut par exemple utiliser pour les coefficients γ_{11} , $\gamma_{21} = \gamma_{12}$ et γ_{22} , les valeurs respectives 1 ; 0,1 et 0,9.

Bien entendu, la variante prévoyant l'utilisation de deux filtres formantiques différents peut être utilisée indépendamment de celle prévoyant l'affaiblissement du gain G_c .

Un tel mode de réalisation est illustré sur la figure 5, où l'on voit que l'utilisation des deux filtres formantiques est prise en combinaison avec l'utilisation du filtre de correction totale.

L'invention s'applique avantageusement à la téléphonie mobile, et en particulier à tous terminaux distants appartenant à un système de communication sans fil.

Un tel terminal, par exemple un téléphone mobile TP, tel que celui illustré sur la figure 6, comporte de façon classique une

antenne reliée par l'intermédiaire d'un duplexeur DUP à une chaîne de réception CHR et à une chaîne de transmission CHT. Un processeur en bande de base BB est relié respectivement à la chaîne de réception CHR et à la chaîne de transmission CHT par l'intermédiaire de convertisseurs analogiques numériques CAN et numériques analogiques CNA.

Classiquement, le processeur BB effectue des traitements en bande de base, et notamment un décodage de canal DCN, suivi d'un décodage de source DCS.

Pour l'émission, le processeur effectue un codage de source CCS suivi d'un codage de canal CCN.

Lorsque le téléphone mobile incorpore un codeur selon l'invention, celui-ci est incorporé au sein des moyens de codage de source CCS, tandis que le décodeur est incorporé au sein des moyens de décodage de source DCS.

REVENDICATIONS

- 1-Procédé d'encodage de la parole à bande élargie, dans lequel on échantillonne la parole de façon à obtenir des trames vocales successives comportant chacune un nombre prédéterminé d'échantillons, et à chaque trame vocale on détermine des paramètres d'un modèle de prédiction linéaire à excitation par code, ces paramètres comportant un mot numérique d'excitation à long terme (v_i) extrait d'un répertoire codé adaptatif (DLT) et un gain à long terme associé (G_a), ainsi qu'un mot d'excitation à court terme (cj) extrait d'un répertoire codé algébrique (DCT) et un gain à court terme associé (G_c), et on met à jour le répertoire codé adaptatif à partir du mot d'excitation à long terme extrait et du mot d'excitation à court terme extrait, caractérisé par le fait qu'on somme (SM) le produit du mot extrait d'excitation à long terme par le gain à long terme associé, avec le produit du mot extrait d'excitation à court terme par le gain à court terme associé, on filtre le mot numérique sommé dans un filtre passe-bas (FLCT) ayant une fréquence de coupure supérieure au quart de la fréquence d'échantillonnage et inférieure à la moitié de celle-ci, et on met à jour le répertoire codé adaptatif avec le mot filtré.
- 2-Procédé selon la revendication 1, caractérisé par le fait qu'on filtre le mot sommé avec un filtre numérique (FLCT) à réponse impulsionnelle finie à phase linéaire ayant un ordre au moins égal à 10.
- 3-Procédé selon la revendication 2, caractérisé par le fait que la fréquence d'échantillonnage est de 16 kHz, et par le fait que le filtre (FLCT) est un filtre d'ordre 20 ayant une fréquence de coupure de l'ordre de 6 kHz.
- 4-Procédé selon l'une des revendications précédentes, caractérisé par le fait que l'extraction du mot d'excitation à court terme comporte un filtrage numérique de prédiction linéaire (FP), et par le fait que le procédé comporte une mise à jour de l'état du filtre de prédiction linéaire avec le mot d'excitation à court terme filtré par un filtre (G_c') dont le ou les coefficients dépendent de la valeur du

gain à long terme, de façon à affaiblir la contribution de l'excitation à court terme lorsque le gain de l'excitation à long terme est supérieur à un seuil prédéterminé.

5 5-Procédé selon la revendication 4, caractérisé par le fait que le seuil prédéterminé est égal à 1.

6-Procédé selon la revendication 5, caractérisé par le fait que le filtre est d'ordre 1 et a une fonction de transfert égale à $B_0 + B_1 z^{-1}$, et par le fait que le premier coefficient du filtre B_0 est égal à $1/(1+\beta.\min(Ga,1))$, et le deuxième coefficient du filtre B_1 est égal à $\beta.\min(Ga,1)/(1+\beta.\min(Ga,1))$, où β est un nombre réel de valeur absolue inférieure à 1, Ga est le gain à long terme et $\min(Ga,1)$ désigne la valeur minimale entre Ga et 1.

7-Procédé selon l'une des revendications précédentes, caractérisé par le fait qu'on effectue l'extraction du mot d'excitation à long terme en utilisant un premier filtre de pondération perceptuelle (FPP1) comportant un premier filtre de pondération formantique, par le fait qu'on effectue l'extraction du mot d'excitation à court terme en utilisant le premier filtre de pondération perceptuelle (FPP1) cascadié à un deuxième filtre de pondération perceptuelle (FPP2) comportant un deuxième filtre de pondération formantique, et par le fait que le dénominateur de la fonction de transfert du premier filtre de pondération formantique est égal au numérateur du deuxième filtre de pondération formantique.

8-Procédé selon la revendication 7 prise en combinaison avec l'une des revendications 4 à 6, caractérisé par le fait qu'il comporte une mise à jour de l'état des deux filtres de pondération perceptuelle avec le mot d'excitation à court terme filtré par ledit filtre d'ordre 1.

9-Dispositif d'encodage de la parole à bande élargie, comportant

- 30 -des moyens d'échantillonnage aptes à échantillonner la parole de façon à obtenir des trames vocales successives comportant chacune un nombre prédéterminé d'échantillons,
- des moyens de traitement aptes à chaque trame vocale, à déterminer des paramètres d'un modèle de prédiction linéaire à

excitation par code, ces moyens de traitement comportant des premiers moyens d'extraction (MEXT1) aptes à extraire un mot numérique d'excitation à long terme d'un répertoire codé adaptatif et à calculer un gain à long terme associé, et des

5 deuxièmes moyens d'extraction (MEXT2) aptes à extraire un mot d'excitation à court terme d'un répertoire codé algébrique et à calculer un gain à court terme associé, et

-des premiers moyens de mise à jour (MAJ) aptes à mettre à jour le répertoire codé adaptatif à partir du mot d'excitation à

10 long terme extrait et du mot d'excitation à court terme extrait, caractérisé par le fait que les premiers moyens de mise à jour comportent

-des premiers moyens de calcul (SM) aptes à sommer le produit du mot extrait d'excitation à long terme par le gain à

15 long terme associé, avec le produit du mot extrait d'excitation à court terme par le gain à court terme associé, de façon à délivrer un mot numérique sommé, et

-un filtre passe-bas (FLCT) ayant une fréquence de coupure supérieure au quart de la fréquence d'échantillonnage et

20 inférieure à la moitié de celle-ci, et connecté entre la sortie des premiers moyens de calcul et le répertoire codé adaptatif de façon à mettre à jour ce répertoire adaptatif avec le mot filtré.

10-Dispositif selon la revendication 9, caractérisé par le fait

25 que le filtre passe-bas (FLCT) est un filtre numérique à réponse impulsionnelle finie à phase linéaire ayant un ordre au moins égal à 10.

11-Dispositif selon la revendication 10, caractérisé par le fait que la fréquence d'échantillonnage est de 16 kHz, et par le fait que le

30 filtre est un filtre d'ordre 20 ayant une fréquence de coupure de l'ordre de 6 kHz.

12-Dispositif selon l'une des revendications 9 à 11, caractérisé par le fait que les premiers moyens d'extraction comportent un filtre numérique de prédiction linéaire (FP), et par le fait que le dispositif comporte des deuxièmes moyens de mise à jour (MAJ2) aptes à effectuer une mise à jour de l'état du filtre de prédiction linéaire avec le mot d'excitation à court terme filtré (G_c') par un filtre dont les coefficients dépendent de la valeur du gain à long terme, de façon à affaiblir la contribution de l'excitation à court terme lorsque le gain de l'excitation à long terme est supérieur à un seuil prédéterminé.

13-Dispositif selon la revendication 12, caractérisé par le fait que le seuil prédéterminé est égal à 1.

14-Dispositif selon la revendication 13, caractérisé par le fait que le filtre est un filtre d'ordre 1 et a une fonction de transfert égale à $B_0 + B_1 z^{-1}$, et par le fait que le premier coefficient B_0 du filtre est égal à $1/(1 + \beta \cdot \min(G_a, 1))$, et le deuxième coefficient B_1 du filtre est égal à $\beta \cdot \min(G_a, 1)/(1 + \beta \cdot \min(G_a, 1))$, où β est un nombre réel de valeur absolue inférieure à 1, G_a est le gain à long terme et $\min(G_a, 1)$ désigne la valeur minimale entre G_a et 1.

15-Dispositif selon l'une des revendications 9 à 14, caractérisé par le fait que les premiers moyens d'extraction comprennent un premier filtre de pondération perceptuelle (FPP1) comportant un premier filtre de pondération formantique, par le fait que les deuxièmes moyens d'extraction comprennent le premier filtre de pondération perceptuelle cascadié à un deuxième filtre de pondération perceptuelle (FPP2) comportant un deuxième filtre de pondération formantique, et par le fait que le dénominateur de la fonction de transfert du premier filtre de pondération formantique est égal au numérateur du deuxième filtre de pondération formantique.

16-Dispositif selon la revendication 15 prise en combinaison avec l'une des revendications 12 à 14, caractérisé par le fait que les deuxièmes moyens de mise à jour sont aptes à effectuer une mise à jour de l'état des deux filtres de pondération perceptuelle avec le mot d'excitation à court terme filtré par ledit filtre d'ordre 1.

17-Terminal d'un système de communication sans fil, caractérisé par le fait qu'il incorpore un dispositif selon l'une des revendications 9 à 16.

5 18-Terminal selon la revendication 17, caractérisé par le fait qu'il forme un téléphone mobile cellulaire.

17. Juli 2002

ABREGE DESCRIPTIF

Procédé et dispositif d'encodage de la parole à bande élargie.

On échantillonne la parole de façon à obtenir des trames vocales successives comportant chacune un nombre prédéterminé d'échantillons. A chaque trame vocale on détermine des paramètres d'un modèle de prédiction linéaire à excitation par code, ces paramètres comportant un mot numérique d'excitation à long terme v_i extrait d'un répertoire codé adaptatif DLT et un gain à long terme associé G_a , ainsi qu'un mot d'excitation à court terme c_j extrait d'un répertoire codé algébrique DCT et un gain à court terme associé G_c . On somme SM le produit du mot extrait d'excitation à long terme par le gain à long terme associé, avec le produit du mot extrait d'excitation à court terme par le gain à court terme associé, on filtre le mot numérique sommé dans un filtre passe-bas FLCT ayant une fréquence de coupure supérieure au quart de la fréquence d'échantillonnage et inférieure à la moitié de celle-ci, et on met à jour le répertoire codé adaptatif avec le mot filtré.

Référence : figure 2.

1/7

FIG. 1
Art antérieur

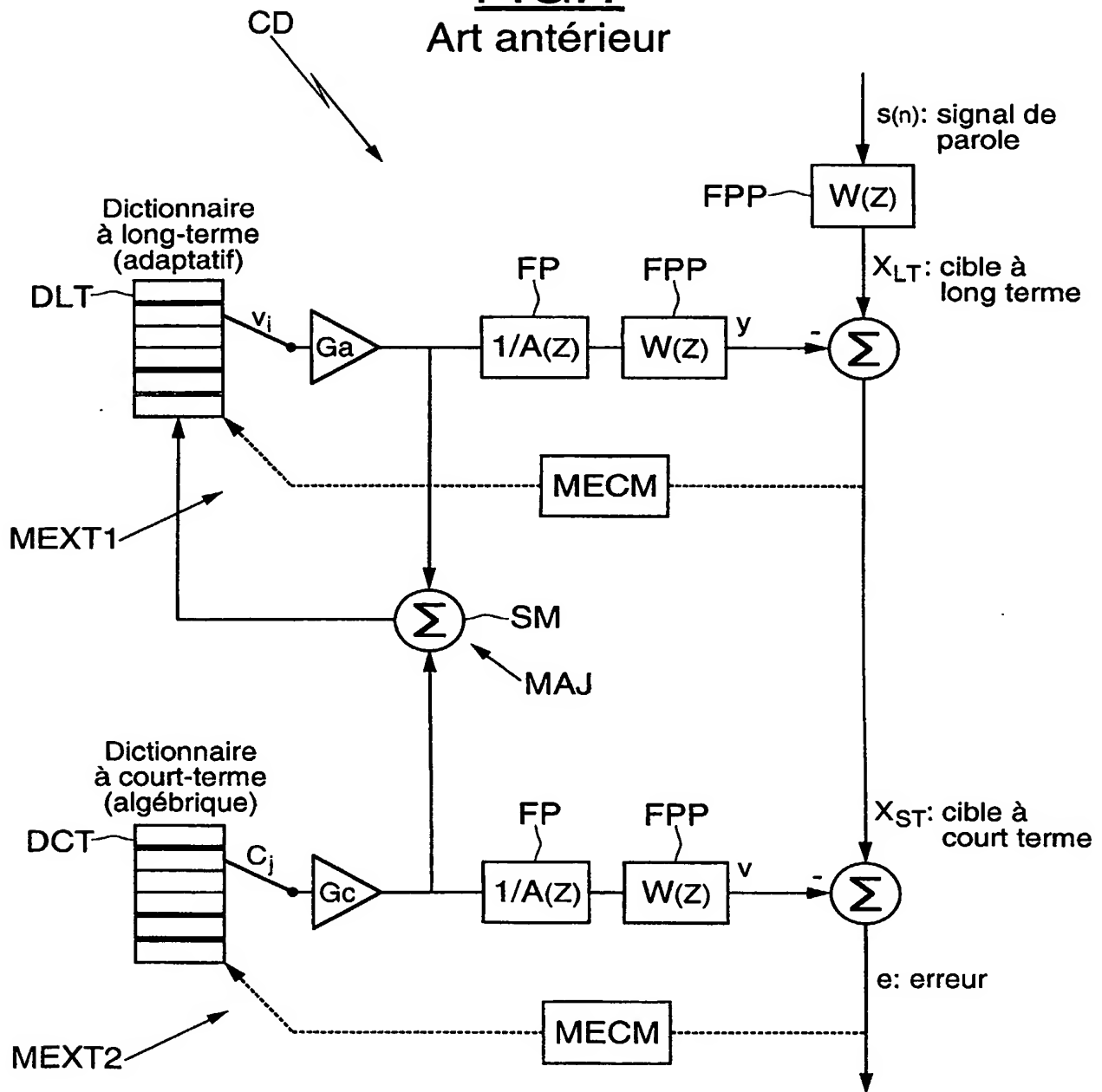


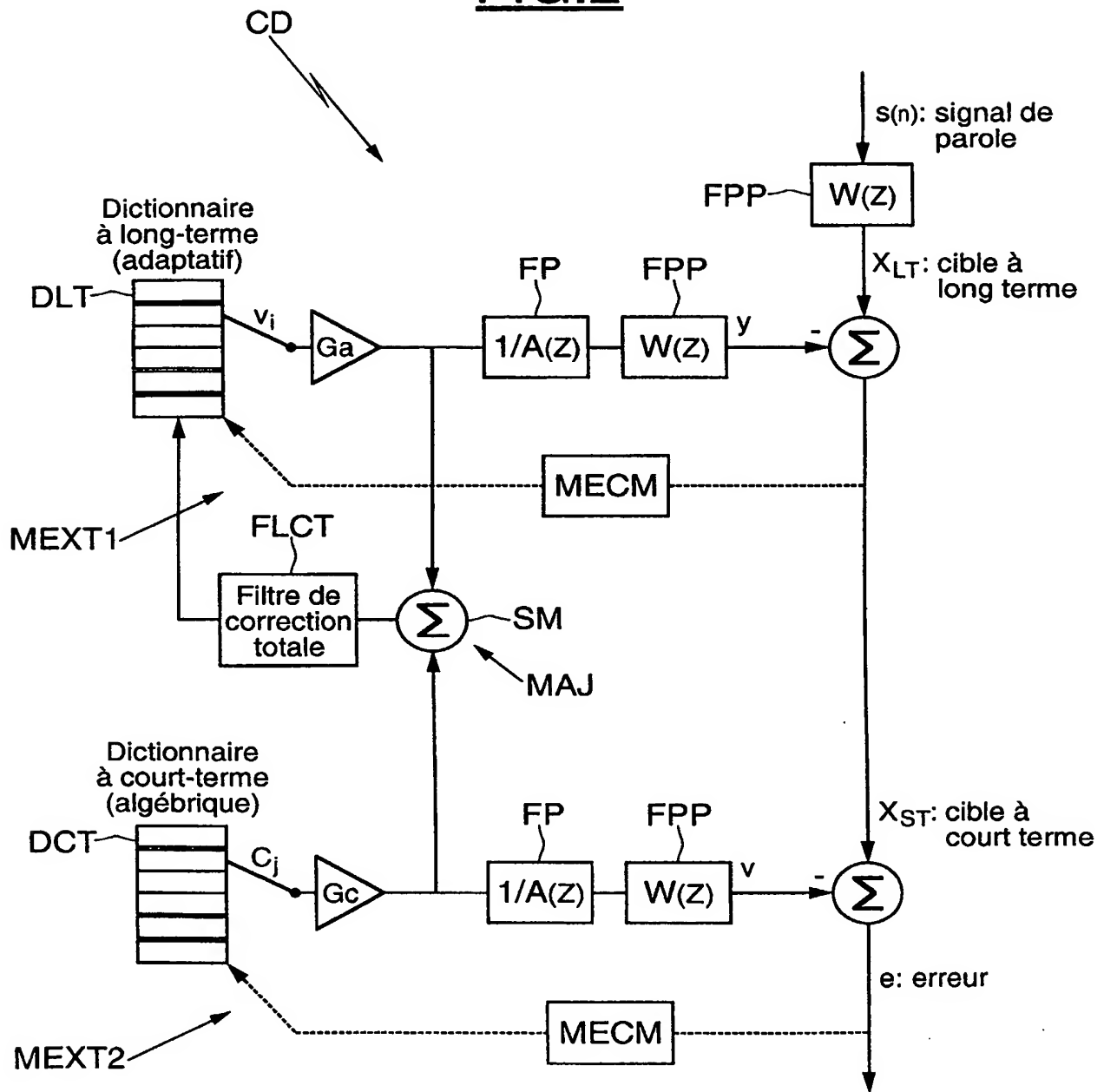
FIG.2

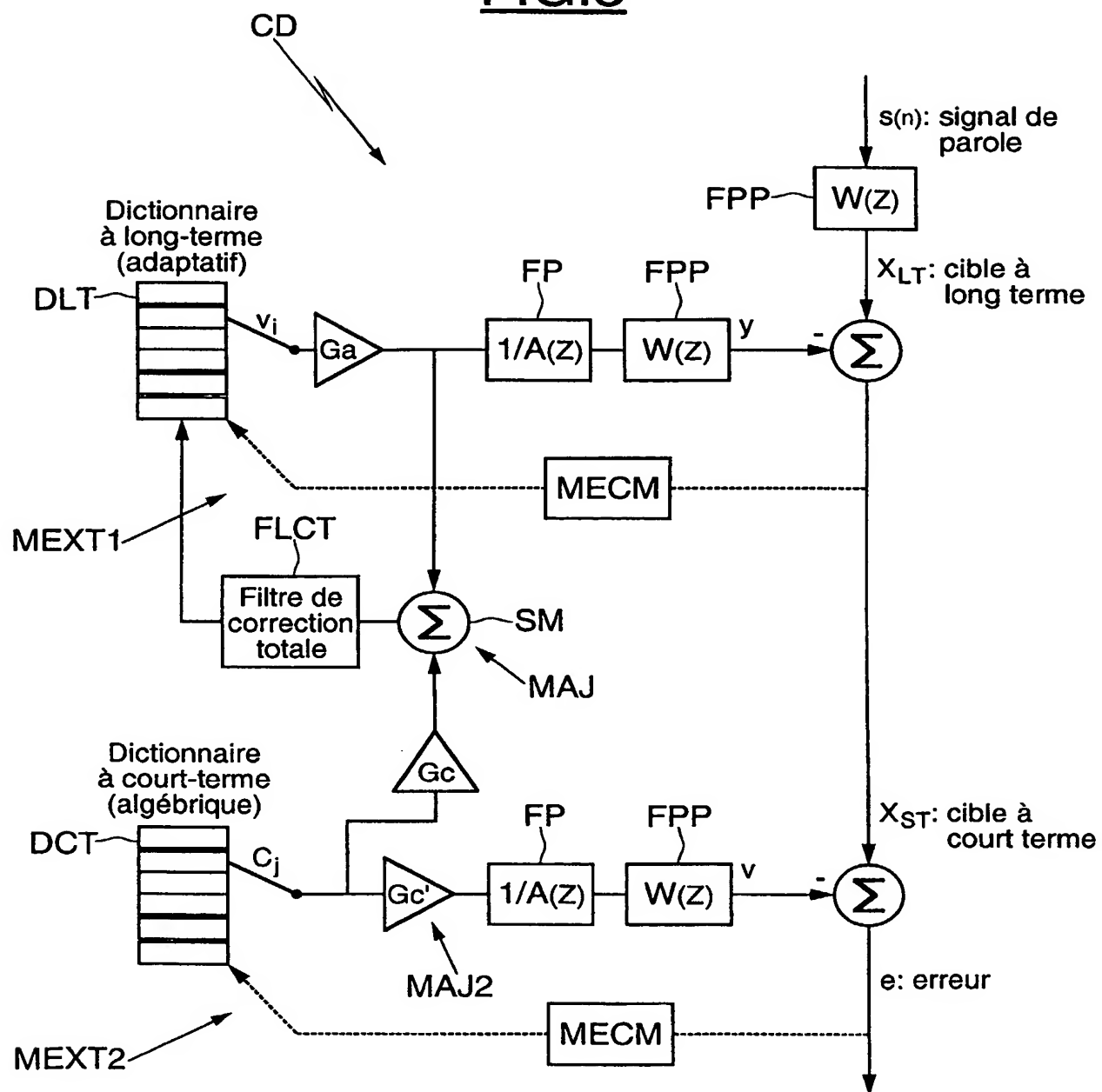
FIG.3

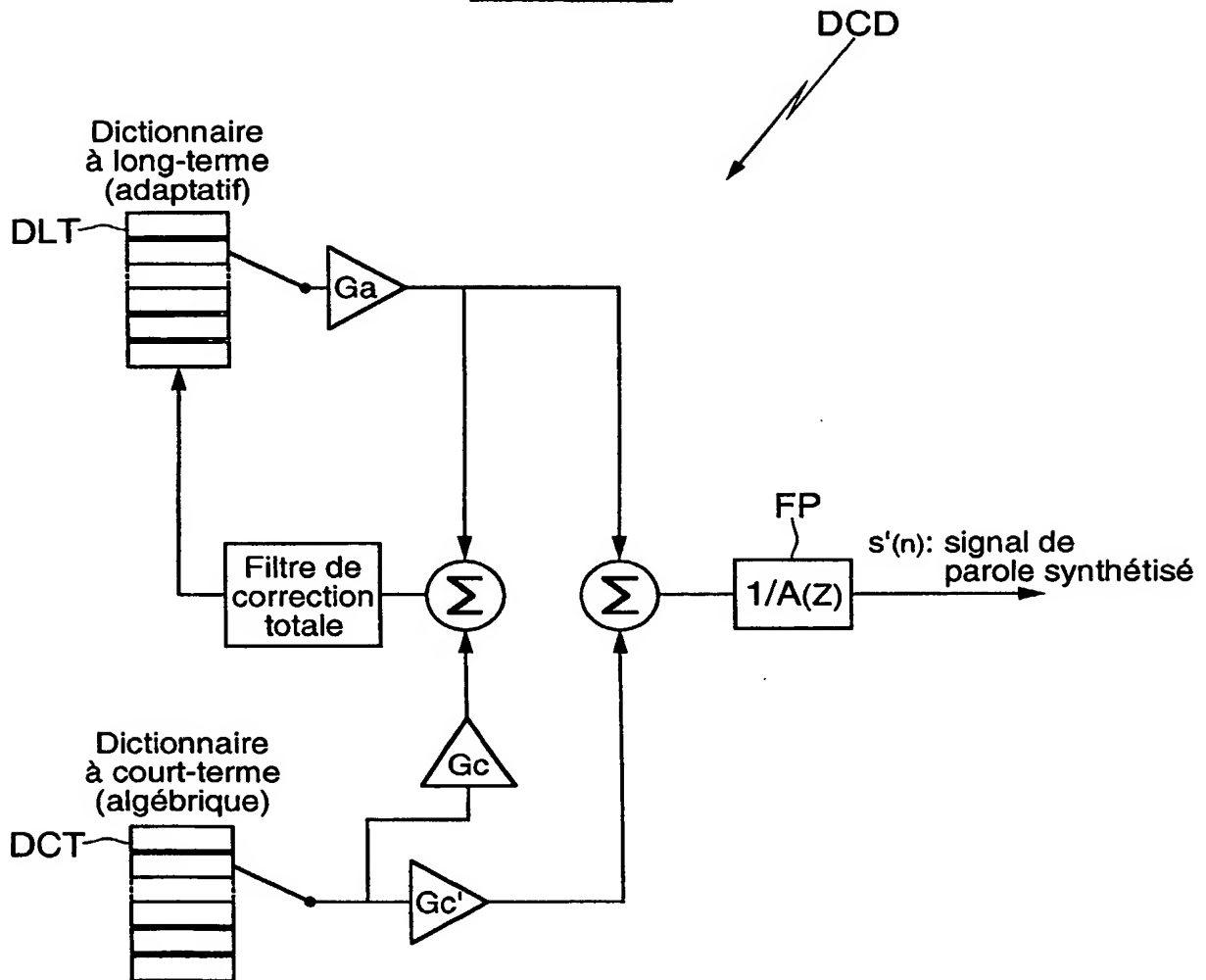
FIG.3a

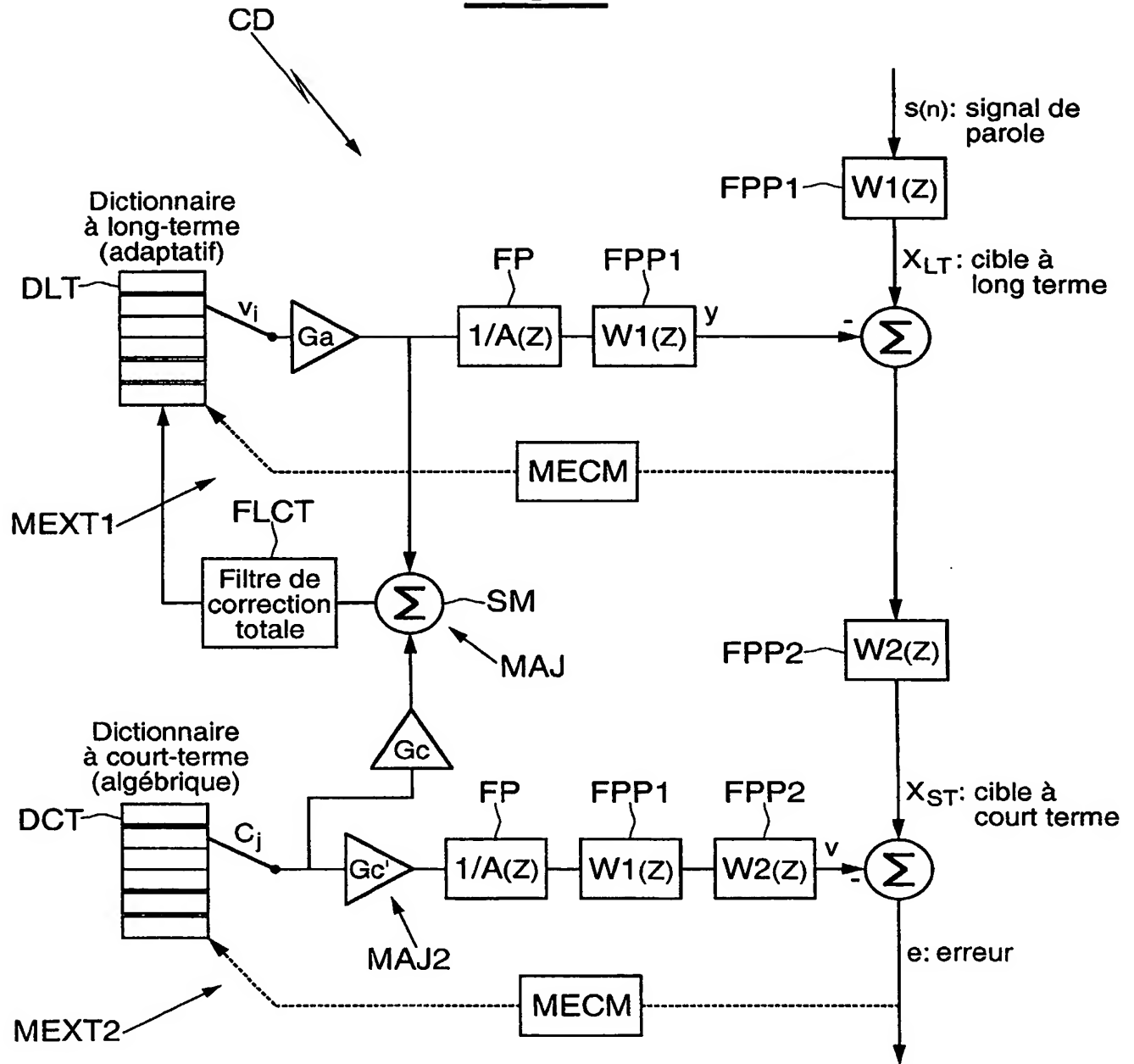
FIG.4

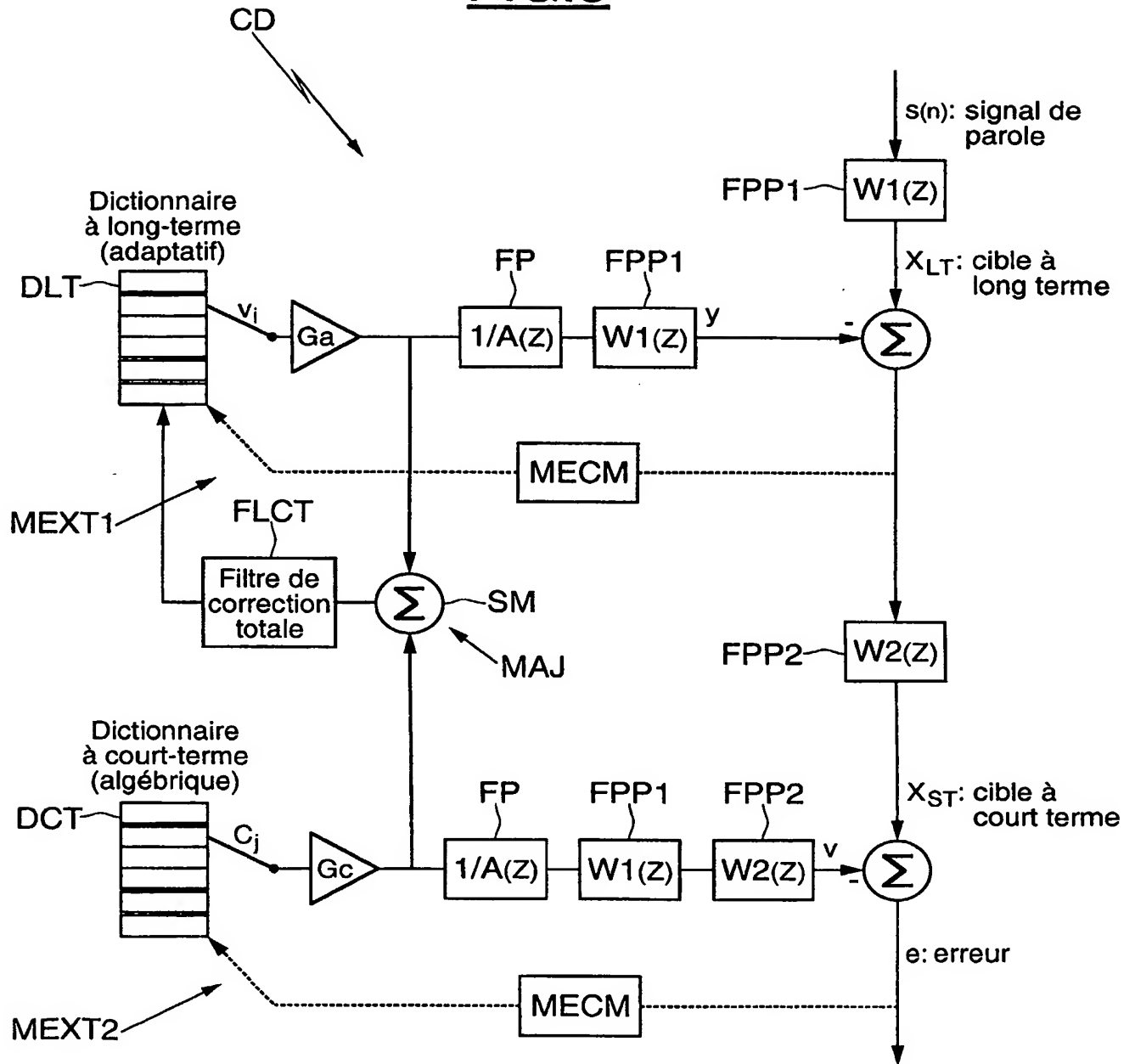
FIG.5

FIG.6